



# A simple proof of the invariant torus theorem

Jacques Fejoz

## ► To cite this version:

Jacques Fejoz. A simple proof of the invariant torus theorem. Regular and Chaotic Dynamics, 2012, 17 (1), pp.1-5. hal-00474037v2

**HAL Id: hal-00474037**

**<https://hal.science/hal-00474037v2>**

Submitted on 14 May 2010

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# A SIMPLE PROOF OF THE INVARIANT TORUS THEOREM

JACQUES FÉJOZ

ABSTRACT. We give a simple proof of Kolmogorov's theorem on the persistence of a quasiperiodic invariant torus in Hamiltonian systems. The theorem is first reduced to a well-posed inversion problem (Herman's normal form) by switching the frequency obstruction from one side of the conjugacy to another. Then the proof consists in applying a simple, well suited, inverse function theorem in the analytic category, which itself relies on the Newton algorithm and on interpolation inequalities. A comparison with other proofs is included in appendix.

## CONTENTS

1. The invariant torus theorem	1
2. Complexification and the functional setting	3
3. Local twisted conjugacy of Hamiltonians	5
Appendices	7
A. An inverse function theorem	7
B. Some estimates on analytic isomorphisms	11
C. Interpolation of spaces of analytic functions	12
D. Weaker arithmetic conditions of convergence	14
E. Comments	15
References	17

## 1. THE INVARIANT TORUS THEOREM

Let  $\mathcal{H}$  be the space of germs along  $T_0^n := \mathbb{T}^n \times \{0\}$  of real analytic Hamiltonians in  $\mathbb{T}^n \times \mathbb{R}^n = \{(\theta, r)\}$  ( $\mathbb{T}^n = \mathbb{R}^n/2\pi\mathbb{Z}^n$ ). The vector field associated with  $H \in \mathcal{H}$  is

$$\vec{H} : \quad \dot{\theta} = \partial_r H, \quad \dot{r} = -\partial_\theta H.$$

For  $\alpha \in \mathbb{R}^n$ , let  $\mathcal{K}$  be the affine subspace of Hamiltonians  $K \in \mathcal{H}$  such that  $K|_{T_0^n}$  is constant (i.e.  $T_0^n$  is invariant) and  $\vec{K}|_{T_0^n} = \alpha$ . Those Hamiltonians are characterized by

---

*Date:* May 2010.

their first order expansion along  $T_0^n$ , of the form  $c + \alpha \cdot r$  for some  $c \in \mathbb{R}$ , that is, their expansion is constant with respect to  $\theta$  and the coefficient of  $r$  is  $\alpha$ .

Let

$$D_{\gamma,\tau} = \{\alpha \in \mathbb{R}^n, \forall k \in \mathbb{Z}^n \setminus \{0\} \mid |k \cdot \alpha| \geq \gamma |k|^{-\tau}\}, \quad |k| := |k|_1 = |k_1| + \cdots + |k_n|.$$

If  $\tau > n - 1$ , the set  $\cup_{\gamma>0} D_{\gamma,\tau}$  has full measure ([Arnold, 1963, p. 83]). See appendix E.

**Theorem 1** (Kolmogorov [1954], Chierchia [2008]). *Let  $\alpha \in D_{\gamma,\tau}$  and  $K^o \in \mathcal{K}$  such that the averaged Hessian*

$$\int_{\mathbb{T}^n} \frac{\partial^2 K^o}{\partial r^2}(\theta, 0) d\theta$$

*is non degenerate. Every  $H \in \mathcal{H}$  close to  $K^o$  possesses an  $\alpha$ -quasiperiodic invariant torus.*

This theorem has far reaching consequences. In particular it has led to a partial answer to the long standing question of the stability of the Solar system (Arnold [1964], Féjoz [2004], Celletti and Chierchia [2007]). See Bost [1986], Sevryuk [2003], de la Llave [2001] for references and background.

Kolmogorov's theorem is a consequence of the following normal form. Let  $\mathcal{G}$  be the space of germs along  $T_0^n$  of real analytic exact symplectomorphisms  $G$  in  $\mathbb{T}^n \times \mathbb{R}^n$  of the following form:

$$G(\theta, r) = (\varphi(\theta), {}^t\varphi'(\theta)^{-1}(r + \rho(\theta))),$$

where  $\varphi$  is a real analytic isomorphism of  $\mathbb{T}^n$  fixing the origin, and  $\rho$  is an exact 1-form on  $\mathbb{T}^n$ .

**Theorem 2** (Herman). *Let  $\alpha \in D_{\gamma,\tau}$  and  $K^o \in \mathcal{K}$ . For every  $H \in \mathcal{H}$  close enough to  $K^o$ , there exists a unique  $(K, G, \beta) \in \mathcal{K} \times \mathcal{G} \times \mathbb{R}^n$  close to  $(K^o, \text{id}, 0)$  such that*

$$H = K \circ G + \beta \cdot r$$

*in some neighborhood of  $G^{-1}(\mathbb{T}_0^n)$ . Moreover,  $\beta$  depends  $C^1$ -smoothly on  $H$ .*

In other words, the orbits of Hamiltonians  $K \in \mathcal{K}$  under the action of symplectomorphisms of  $\mathcal{G}$  locally form a subspace of finite codimension  $n$ . The offset  $\beta \cdot r$  usually breaks the dynamical conjugacy between  $K$  and  $H$ ; hence Herman's normal form is of geometrical nature and can be called a *twisted conjugacy*. The strategy for deducing the existence of an  $H$ -invariant torus (namely,  $G^{-1}(\mathbb{T}_0^n)$ ) from that of a  $K$ -invariant torus (namely,  $\mathbb{T}_0^n$ ) is to show that  $\beta$  vanishes on some subset of large measure in some parameter space (in some cases, the frequency  $\alpha$  cannot be fixed and needs to be varied).

In the paper,  $O(r^n)$  will denote the ideal of functions of  $(\theta, r)$  of the  $n$ -th order with respect to  $r$ .

*Proof of theorem 1 assuming theorem 2.* Let  $K_2^o(\theta) := \frac{1}{2} \frac{\partial^2 K^o}{\partial r^2}(\theta, 0)$ . Let  $F$  be the analytic function taking values among symmetric bilinear forms, which solves the cohomological equation  $\mathcal{L}_\alpha F = K_2^o - \int_{\mathbb{T}_0^n} K_2^o d\theta$  (see lemma 5), and  $\varphi$  be the germ along  $\mathbb{T}_0^n$  of the (well defined) time-one map of the flow of the Hamiltonian  $F(\theta) \cdot r^2$ . The map  $\varphi$  is

symplectic and restricts to the identity on  $T_0^n$ . At the expense of substituting  $K^o \circ \varphi$  and  $H \circ \varphi$  for  $K^o$  and  $H$  respectively, one can thus assume that

$$K^o = c + \alpha \cdot r + Q \cdot r^2 + O(r^3), \quad Q := \int_{T_0^n} K_2^o(\theta) d\theta.$$

The germs so obtained from the initial  $K^o$  and  $H$  are close to one another.

Consider the family of trivial perturbations obtained by translating  $K^o$  in the direction of actions:

$$K_R^o(\theta, r) := K^o(\theta, R + r), \quad R \in \mathbb{R}^n, \quad R \text{ small},$$

and its approximation obtained by truncating the first order jet of  $K_R^o$  along  $T_0^n$  from its terms  $O(R^2)$  which possibly depend on  $\theta$ :

$$\hat{K}_R^o(\theta, r) := (c + \alpha \cdot R) + (\alpha + 2Q \cdot R) \cdot r + O(r^2) = K_R^o + O(R^2).$$

For the Hamiltonian  $\hat{K}_R^o$ ,  $T_0^n$  is invariant and quasiperiodic of frequency  $\alpha + 2Q \cdot R$ . Hence the Herman normal form of  $\hat{K}_R^o$  with respect to the frequency  $\alpha$  is

$$\hat{K}_R^o = \left( \hat{K}_R^o - \hat{\beta}_R^o \cdot r \right) \circ \text{id} + \hat{\beta}_R^o \cdot r, \quad \hat{\beta}_R^o := 2Q \cdot R.$$

By assumption the matrix  $\left. \frac{\partial \hat{\beta}^o}{\partial R} \right|_{R=0} = 2Q$  is invertible and the map  $R \mapsto \hat{\beta}^o(R)$  is a local diffeomorphism.

Now, theorem 2 asserts the existence of an analogous map  $R \mapsto \beta(R)$  for  $H_R$ , which is a small  $C^1$ -perturbation of  $R \mapsto \hat{\beta}^o(R)$ , and thus a local diffeomorphism, with a domain having a lower bound locally uniform with respect to  $H$ . Hence if  $H$  is close enough to  $K^o$  there is a unique small  $R$  such that  $\beta = 0$ . For this  $R$  the equality  $H_R = K \circ G$  holds, hence the torus obtained by translating  $G^{-1}(T_0^n)$  by  $R$  in the direction of actions is invariant and  $\alpha$ -quasiperiodic for  $H$ .  $\square$

**Exercise 3** Simplify this proof when  $K^o = K^o(r)$  is integrable.

It is the aim of the rest of the paper to prove theorem 2, by locally inverting some operator

$$\phi : (K, G, \beta) \mapsto H = K \circ G + \beta \cdot r$$

when  $\alpha$  is diophantine.

## 2. COMPLEXIFICATION AND THE FUNCTIONAL SETTING

For various sets  $U$  and  $V$ ,  $\mathcal{A}(U, V)$  will denote the set of continuous maps  $U \rightarrow V$  which are real analytic on the interior  $\overset{\circ}{U}$ , and  $\mathcal{A}(U) := \mathcal{A}(U, \mathbb{C})$ .

Recall notations for the abstract torus and its embedding in the phase space:

$$\mathbb{T}^n = \mathbb{R}^n / 2\pi\mathbb{Z}^n \quad \text{and} \quad T_0^n = \mathbb{T}^n \times \{0\} \subset \mathbb{T}^n \times \mathbb{R}^n.$$

Define complex extensions

$$\mathbb{T}_{\mathbb{C}}^n = \mathbb{C}^n / 2\pi\mathbb{Z}^n \quad \text{and} \quad T_{\mathbb{C}}^n = \mathbb{T}_{\mathbb{C}}^n \times \mathbb{C}^n$$

as well as bases of neighborhoods

$$\mathbb{T}_s^n = \{\theta \in \mathbb{T}_{\mathbb{C}}^n, \max_{1 \leq j \leq n} |\operatorname{Im} \theta_j| \leq s\} \quad \text{and} \quad \mathbb{T}_s^n = \{(\theta, r) \in \mathbb{T}_{\mathbb{C}}^n, |(\theta, r)| \leq s\},$$

with  $|(\theta, r)| := \max_{1 \leq j \leq n} \max(|\operatorname{Im} \theta_j|, |r_j|)$ .

**2.1. Spaces of Hamiltonians.** – Let  $\mathcal{H}_s = \mathcal{A}(\mathbb{T}_s^n)$ , endowed with the Banach norm

$$|H|_s := \sup_{(\theta, r) \in \mathbb{T}_s^n} |H(\theta, r)|,$$

so that  $\mathcal{H}$  be the inductive limit of the spaces  $\mathcal{H}_s$ .

– For  $\alpha \in \mathbb{R}^n$ , let  $\mathcal{K}_s$  be the affine subspace consisting of those  $K \in \mathcal{H}_s$  such that  $K(\theta, r) = c + \alpha \cdot r + O(r^2)$  for some  $c \in \mathbb{R}$ .

– If  $G$  is a real analytic isomorphism on some open set of  $\mathbb{T}_{\mathbb{C}}^n$  and if  $G$  is transverse to  $\mathbb{T}_s^n$ , let  $G^* \mathcal{A}(\mathbb{T}_s^n) := \mathcal{A}(G^{-1}(\mathbb{T}_s^n))$  be endowed with the Banach norm

$$|H|_{G,s} := |H \circ G^{-1}|_s.$$

**2.2. Spaces of conjugacies.**

**2.2.1. Diffeomorphisms of the torus.** Let  $\mathcal{D}_s$  be the space of maps  $\varphi \in \mathcal{A}(\mathbb{T}_s^n, \mathbb{T}_{\mathbb{C}}^n)$  which are analytic isomorphisms from  $\mathring{\mathbb{T}}_s^n$  to their image and which fix the origin.

Let also

$$\chi_s := \{v \in \mathcal{A}(\mathbb{T}_s^n)^n, v(0) = 0\}$$

be the space of vector fields on  $\mathbb{T}_s^n$  which vanish at 0, endowed with the Banach norm

$$|v|_s := \max_{\theta \in \mathbb{T}_s^n} \max_{1 \leq j \leq n} |v_j(\theta)|.$$

According to corollary 14, the map

$$\sigma \mathring{B}_{s+\sigma}^\chi := \{v \in \chi_{s+\sigma}, |v|_s < \sigma\} \rightarrow \mathcal{D}_s, \quad v \mapsto \operatorname{id} + v$$

is defined and locally bijective. It endows  $\mathcal{D}_s$  with a local structure of Banach manifold in the neighborhood of the identity.

We will consider the contragredient action of  $\mathcal{D}_s$  on  $\mathbb{T}_s^n$  (with values in  $\mathbb{T}_{\mathbb{C}}^n$ ) :

$$\varphi(\theta, r) := (\varphi(\theta), {}^t \varphi'(\theta)^{-1} \cdot r),$$

in order to linearize the dynamics on the alleged invariant tori.

**2.2.2. Straightening tori.** Let  $\mathcal{B}_s$  be the space of exact one-forms over  $\mathbb{T}_s^n$ , with

$$|\rho|_s = \max_{\theta \in \mathbb{T}_s^n} \max_{1 \leq j \leq n} |\rho_j(\theta)|, \quad \rho = (\rho_1, \dots, \rho_n).$$

We will consider its action on  $\mathbb{T}_s^n$  by translation of the actions:

$$\rho(\theta, r) := (\theta, r + \rho(\theta)),$$

in order to straighten the perturbed invariant tori.

2.2.3. *Our space of conjugacies.* Let  $\mathcal{G}_s = \mathcal{D}_s \times \mathcal{B}_s$ , identified with a space of Hamiltonian symplectomorphisms by

$$(\varphi, \rho)(\theta, r) := \varphi \circ \rho(\theta, r) = (\varphi(\theta), {}^t\varphi'(\theta)^{-1}(r + \rho(\theta))).$$

Endow its tangent space at the identity  $T_{\text{id}}\mathcal{G}_s = \mathfrak{g}_s := \chi_s \times \mathcal{B}_s$  with the norm

$$|\dot{G}|_s = |(v, \rho)|_s := \max(|v|_s, |\rho|_s),$$

and its tangent space at  $G = (\varphi, \rho)$  with the norm

$$|\delta G|_s := |\delta G \circ G^{-1}|_s, \quad \delta G \in T_G\mathcal{G}.$$

Here and elsewhere, the notation  $\delta G$ , as well as similar ones, should be taken as a whole; there is no separate  $\delta \in \mathbb{R}$  in the present paper.

Also consider the following neighborhoods of the identity:

$$\mathcal{G}_s^\sigma = \left\{ G \in \mathcal{G}_s, \max_{(\theta, r) \in T_s^n} |(\Theta - \theta, R - r)| \leq \sigma, (\Theta, R) = G(\theta, r) \right\}, \quad \sigma > 0.$$

The operators (commuting with inclusions of source and target spaces)

$$\phi_s : E_s := \mathcal{K}_{s+\sigma} \times \mathcal{G}_s^\sigma \times \mathbb{R}^n \rightarrow \mathcal{H}_s, \quad (K, G, \beta) \mapsto K \circ G + \beta \cdot r$$

are now defined.

### 3. LOCAL TWISTED CONJUGACY OF HAMILTONIANS

**Theorem 4.** *Let  $\alpha \in D_{\gamma, \tau}$ . For all  $0 < s < s + \sigma < 1$ ,  $\phi_{s+\sigma}$  has a local inverse: if  $|H - K^o|_{s+\sigma}$  is small, there is a unique  $(K, G, \beta) \in E_s$ ,  $|\cdot|_s$ -close to  $(K^o, \text{id}, 0)$  such that  $H = K \circ G + \beta \cdot r$ . Moreover  $\beta \circ \phi^{-1}$  is a  $C^1$ -function locally in the neighborhood of  $K^o$  in  $\mathcal{H}_{s+\sigma}$ .*

This entails theorem 2 and itself follows from the inverse function theorem of appendix A, from lemma 11 (for the uniqueness) and from corollary 13 (for the smoothness of  $\beta \circ \phi^{-1}$ ). We will now check the two main hypotheses of appendix A (one on  $\phi'^{-1}$  and one on  $\phi''$ ).

Let  $\mathcal{L}_\alpha$  be the Lie derivative operator in the direction of the constant vector field  $\alpha$ :

$$\mathcal{L}_\alpha : \mathcal{A}(\mathbb{T}_s^n) \rightarrow \mathcal{A}(\mathbb{T}_s^n), \quad f \mapsto f' \cdot \alpha = \sum_{1 \leq j \leq n} \alpha_j \frac{\partial f}{\partial \theta_j}.$$

We will need the following classical lemma in two instances in the proof of lemma 6.

**Lemma 5** (Cohomological equation). *If  $g \in \mathcal{A}(\mathbb{T}_{s+\sigma}^n)$  has 0-average ( $\int_{\mathbb{T}} g d\theta = 0$ ), there exists a unique function  $f \in \mathcal{A}(\mathbb{T}_s^n)$  of 0-average such that  $\mathcal{L}_\alpha f = g$ , and there exists a  $C_0 = C_0(n, \tau)$  such that, for any  $\sigma$ :*

$$|f|_s \leq C_0 \gamma^{-1} \sigma^{-\tau-n} |g|_{s+\sigma}.$$

*Proof.* Let  $g(\theta) = \sum_{k \in \mathbb{Z}^n \setminus \{0\}} g_k e^{ik \cdot \theta}$  be the Fourier expansion of  $g$ . The unique formal solution to the equation  $\mathcal{L}_\alpha f = g$  is given by  $f(\theta) = \sum_{k \in \mathbb{Z}^n \setminus \{0\}} \frac{g_k}{ik \cdot \alpha} e^{ik \cdot \theta}$ .

Since  $g$  is analytic, its Fourier coefficients decay exponentially: we find

$$|g_k| = \left| \int_{\mathbb{T}^n} g(\theta) e^{-ik \cdot \theta} \frac{d\theta}{2\pi} \right| \leq |g|_{s+\sigma} e^{-|k|(s+\sigma)}$$

by shifting the torus of integration to a torus  $\text{Im } \theta_j = \pm(s + \sigma)$ .

Using this estimate and replacing the small denominators  $k \cdot \alpha$  by the estimate defining the diophantine property of  $\alpha$ , we get

$$\begin{aligned} |f|_s &\leq \frac{|g|_{s+\sigma}}{\gamma} \sum_k |k|^\tau e^{-|k|\sigma} \\ &\leq \frac{2^n |g|_{s+\sigma}}{\gamma} \sum_{\ell \geq 1} \binom{\ell+n-1}{\ell} \ell^\tau e^{-\ell\sigma} \leq \frac{4^n |g|_{s+\sigma}}{\gamma (n-1)!} \sum_{\ell} (\ell+n-1)^{\tau+n-1} e^{-\ell\sigma}, \end{aligned}$$

where the latter sum is bounded by

$$\begin{aligned} \int_1^\infty (\ell+n-1)^{\tau+n-1} e^{-(\ell-1)\sigma} d\ell &= \sigma^{-\tau-n} e^{n\sigma} \int_{n\sigma}^\infty \ell^{\tau+n-1} e^{-\ell} d\ell \\ &< \sigma^{-\tau-n} e^{n\sigma} \int_0^\infty \ell^{\tau+n-1} e^{-\ell} d\ell = \sigma^{-\tau-n} e^{n\sigma} \Gamma(\tau+n). \end{aligned}$$

Hence  $f$  belongs to  $\mathcal{A}(\mathbb{T}_s^n)$  and satisfies the wanted estimate.  $\square$

We will write  $x = (K, G, \beta, c)$ ,  $\delta x = (\delta K, \delta G, \delta \beta, \delta c)$  and  $\delta \hat{x} = (\delta \hat{K}, \delta \hat{G}, \delta \hat{\beta}, \delta \hat{c})$ .

Fix  $0 < s < s + \sigma < 1$ .

**Lemma 6.** *There exists  $C' > 0$  which is locally uniform with respect to  $x \in E_s$  in the neighborhood of  $G = \text{id}$  such that the linear map  $\phi'(x)$  has an inverse  $\phi'(x)^{-1}$  satisfying*

$$|\phi'(x)^{-1} \cdot \delta H|_s \leq \sigma^{-\tau-n-1} C' |\delta H|_{G, s+\sigma}.$$

*Proof.* A function  $\delta H \in G^* \mathcal{A}(T_{s+\sigma})$  being given, we want to solve the equation

$$\delta \phi(x) \cdot \delta x = \delta K \circ G + K' \circ G \cdot \delta G + \delta \beta \cdot r + \delta c = \delta H,$$

for the unknowns  $\delta K \in T_K \mathcal{K}_s \subset \mathcal{A}(T_s^n)$ ,  $\delta G \in T_G \mathcal{G}_s$ ,  $\delta \beta \in \mathbb{R}^n$  and  $\delta c \in \mathbb{R}$ , or, equivalently, after composing with  $G^{-1}$  to the right,

$$\delta K + K' \cdot \dot{G} + \delta \beta \cdot r \circ G^{-1} + \delta c = \dot{H},$$

where we have set  $\dot{G} := \delta G \circ G^{-1} \in \mathfrak{g}_s$  and  $\dot{H} := \delta H \circ G^{-1} \in \mathcal{A}(T_s^n)$ .

More specifically,  $G^{-1}$  and  $\dot{G}$  are of the form

$$G^{-1}(\theta, r) = (\varphi^{-1}(\theta), {}^t \varphi' \circ \varphi^{-1}(\theta) \cdot r - \rho \circ \varphi^{-1}(\theta)), \quad \dot{G} = (\dot{\varphi}, \dot{\rho} - r \cdot \dot{\varphi}'),$$

where  $\dot{\varphi} \in \chi_{s+\sigma}$  and  $\dot{\rho} \in \mathcal{B}_{s+\sigma}$ , and we can expand

$$K = \alpha \cdot r + K_2(\theta) \cdot r^2 + O(r^3) \quad \text{and} \quad \dot{H} = \dot{H}_0(\theta) + \dot{H}_1(\theta) \cdot r + O(r^2).$$

The equation becomes

$$(1) \quad [\dot{\rho} \cdot \alpha + \delta c - \rho \circ \varphi^{-1} \cdot \delta \beta] + r \cdot [-\dot{\varphi}' \cdot \alpha + \varphi' \circ \varphi^{-1} \cdot \delta \beta + 2K_2 \cdot \dot{\rho}] + \dot{K} = \dot{H} + O(r^2),$$

where the term  $O(r^2)$  in the right hand side depends only on  $K$  and  $\dot{G}$ , and not on  $\dot{K}$ . The equation turns out to be triangular in the five unknowns. The existence and uniqueness of a solution with the wanted estimate follows from repeated applications of lemma 5 and Cauchy's inequality:

– The average over  $T_0^n$  of the first order terms with respect to  $r$  in equation (1) yields

$$\delta \beta = \left( \int_{T_0^n} \varphi' \circ \varphi^{-1} d\theta \right)^{-1} \cdot \int_{T_0^n} \dot{H}_1 d\theta,$$

which does exist if  $\varphi$  is close to the identity (proposition 14).

– Similarly, the average of the restriction to  $T_0^n$  of (1) yields:

$$\delta c = \int_{T_0^n} \dot{H}_0 d\theta + \int_{T_0^n} \rho \circ \varphi^{-1} d\theta \cdot \delta \beta.$$

– Next, the restriction to  $T_0^n$  of (1) can be solved uniquely with respect to  $\delta \rho$  according to lemma 5 (applied with  $\rho = f'$ ).

– The part of degree one can then be solved for  $\dot{\varphi}$  similarly.

– Terms of order  $\geq 2$  in  $r$  determine  $\dot{K}$ . □

**Lemma 7.** *There exists a constant  $C'' > 0$  which is locally uniform with respect to  $x \in E_{s+\sigma}$  in the neighborhood of  $G = \text{id}$  such that the bilinear map  $\phi''(x)$  satisfies*

$$|\phi''(x) \cdot \delta x \otimes \delta \hat{x}|_{G,s} \leq \sigma^{-1} C'' |\delta x|_{s+\sigma} |\delta \hat{x}|_{s+\sigma}.$$

*Proof.* Differentiating  $\phi$  twice yields

$$\phi''(x) \cdot \delta x \otimes \delta \hat{x} = \delta K' \circ G \cdot \delta G + \delta \hat{K}' \circ G \cdot \delta G + K'' \circ G \cdot \delta G \otimes \delta \hat{G},$$

whence the estimate. □

## A. AN INVERSE FUNCTION THEOREM

Let  $E = (E_s)_{0 < s < 1}$  be a decreasing family of Banach spaces with increasing norms  $|\cdot|_s$ , and  $\epsilon B_s^E = \{x \in E_s, |x|_s < \epsilon\}$ ,  $\epsilon > 0$ , be its balls centered at 0.

Let  $(F_s)$  be an analogous family. Endow  $F$  with additional norms  $|\cdot|_{x,s}$ ,  $x \in E_s$ ,  $0 < s < 1$ , satisfying

$$|y|_{0,s} = |y|_s \quad \text{and} \quad |y|_{x',s} \leq |y|_{x,s+|x'-x|_s}.$$

These norms allow for dealing with composition operators without artificially loosing some fixed “width of analyticity”  $\sigma$  at each step of the Newton algorithm.



Let  $\phi : \sigma B_{s+\sigma}^E \rightarrow F_s$ ,  $s < s + \sigma$ ,  $\phi(0) = 0$ , be maps commuting with inclusions, twice differentiable, such that the differential  $\phi'(x) : E_{s+\sigma} \rightarrow F_s$  has a right inverse  $\phi'(x)^{-1} : F_{s+\sigma} \rightarrow E_s$ , and

$$\begin{cases} |\phi'(x)^{-1}\eta|_s \leq C'\sigma^{-\tau'}|\eta|_{x,s+\sigma} \\ |\phi''(x)\xi^{\otimes 2}|_{x,s} \leq C''\sigma^{-\tau''}|\xi|_{s+\sigma}^2 \end{cases} \quad (\forall s, \sigma, x, \xi, \eta)$$

with  $C', C'', \tau', \tau'' \geq 1$ . Let  $C := C'C''$  and  $\tau := \tau' + \tau''$ .

**Theorem 8.**  *$\phi$  is locally surjective and, more precisely, for any  $s, \eta$  and  $\sigma$  with  $\eta < s$ ,*

$$\epsilon B_{s+\sigma}^F \subset \phi(\eta B_s^E), \quad \epsilon := 2^{-8\tau} C^{-2} \sigma^{2\tau} \eta.$$

In other words,  $\phi$  has a right-inverse  $\psi : \epsilon B_{s+\sigma}^F \rightarrow \eta B_s^E$ .

*Proof.* Some numbers  $s, \eta$  and  $\sigma$  and  $y \in B_{s+\eta}^F$  being given, let

$$f : \sigma B_{s+\eta+\sigma}^E \rightarrow E_s, \quad x \mapsto x + \phi'(x)^{-1}(y - \phi(x))$$

and

$$Q : \sigma B_{s+\sigma}^E \times \sigma B_{s+\sigma}^E \rightarrow F_s, \quad (x, \hat{x}) \mapsto \phi(\hat{x}) - \phi(x) - \phi'(x)(\hat{x} - x).$$

**Lemma 9.** *The function  $Q$  satisfies:  $|Q(x, \hat{x})|_{x,s} \leq 2^{-1} C'' \sigma^{-\tau''} |\hat{x} - x|_{s+\sigma+|\hat{x}-x|_s}^2$ .*

*Proof of the lemma.* Let  $\hat{x}_t := (1-t)x + t\hat{x}$ . Taylor's formula yields

$$Q(x, \hat{x}) = \int_0^1 (1-t) \phi''(\hat{x}_t) (\hat{x} - x)^2 dt,$$

hence

$$|Q(x, \hat{x})|_{x,s} \leq \int_0^1 (1-t) |\phi''(\hat{x}_t)(\hat{x} - x)^2|_{x,s} dt \leq \int_0^1 (1-t) |\phi''(\hat{x}_t)(\hat{x} - x)^2|_{\hat{x}_t, s+|\hat{x}_t-x|_s} dt,$$

whence the estimate.  $\square$

Now, let  $s, \eta$  and  $\sigma$  be fixed, with  $\eta < s$  and  $y \in \epsilon B_{s+\sigma}^F$  for some  $\epsilon$ . We will see that if  $\epsilon$  is small enough, the sequence  $x_0 = 0$ ,  $x_n := f^n(0)$  is defined for all  $n \geq 0$  and converges towards some preimage  $x \in \eta B_s^E$  of  $y$  by  $\phi$ .

Let  $(\sigma_n)_{n \geq 0}$  be a sequence of positive real numbers such that  $3 \sum \sigma_n = \sigma$ , and  $(s_n)_{n \geq 0}$  be the sequence decreasing from  $s_0 := s + \sigma$  to  $s$  defined by induction by the formula  $s_{n+1} = s_n - 3\sigma_n$ .

Assuming the existence of  $x_0, \dots, x_{n+1}$ , we see that  $\phi(x_k) = y + Q(x_{k-1}, x_k)$ , hence

$$x_{k+1} - x_k = \phi'(x_k)^{-1}(y - \phi(x_k)) = -\phi'(x_k)^{-1}Q(x_{k-1}, x_k) \quad (1 \leq k \leq n).$$

Further assuming that  $|x_{k+1} - x_k|_{s_k} \leq \sigma_k$ , the estimate of the right inverse and lemma 9 entail that

$$|x_{n+1} - x_n|_{s_{n+1}} \leq c_n |x_n - x_{n-1}|_{s_n}^2 \leq \dots \leq c_n c_{n-1}^2 \dots c_1^{2^{n-1}} |x_1|_{s_1}^{2^{n-1}}, \quad c_k := 2^{-1} C \sigma_k^{-\tau}.$$

The estimate

$$|x_1|_{s_1} \leq C'(3\sigma_0)^{-\tau'} |y|_{s_0} \leq 2^{-1} C \sigma_0^{-\tau} \epsilon = c_0 \epsilon$$

and the fact, to be checked later, that  $c_k \geq 1$  for all  $k \geq 0$ , show :

$$|x_{n+1} - x_n|_{s_{n+1}} \leq \left( \epsilon \prod_{k \geq 0} c_k^{2^{-k}} \right)^{2^n}.$$

Since  $\sum_{n \geq 0} \rho^{2^n} \leq 2\rho$  if  $2\rho \leq 1$ , and using the definition of constants  $c_k$ 's, we get a sufficient condition to have all  $x_n$ 's defined and to have  $\sum |x_{n+1} - x_n|_s \leq \eta$ :

$$(2) \quad \epsilon = \frac{\eta}{2} \prod_{k \geq 0} c_k^{-2^{-k}} = \frac{2\eta}{C^2} \prod_{k \geq 0} \sigma_k^{\tau 2^{-k}}.$$

Maximizing the upper bound of  $\epsilon$  under the constraint  $3 \sum_{n \geq 0} \sigma_n = \sigma$  yields  $\sigma_k := \frac{\sigma}{6} 2^{-k}$ . A posteriori it is straightforward that  $|x_{n+1} - x_n|_{s_n} \leq \sigma_n$  (as earlier assumed to apply lemma 9) and  $c_n \geq 1$  for all  $n \geq 0$ . Besides, using that  $\sum k 2^{-k} = \sum 2^{-k} = 2$  we get

$$\frac{\eta}{2} \prod_{k \geq 0} c_k^{-2^{-k}} = \frac{\eta}{2} \prod_{k \geq 0} \frac{1}{2^{\tau k 2^{-k}}} \left( \frac{2}{C} \left( \frac{\sigma}{6} \right)^\tau \right)^{2^{-k}} = \frac{2\eta}{C^2} \left( \frac{\sigma}{12} \right)^{2\tau} > \frac{\sigma^{2\tau} \eta}{2^{8\tau} C^2},$$

whence the theorem.  $\square$

**Exercise 10** The domain of  $\psi$  contains  $\epsilon B_S^F$ ,  $\epsilon = 2^{-12\tau} \tau^{-1} C^{-2} S^{3\tau}$ , for any  $S$ .

*Proof.* The above function  $\epsilon(\eta, \sigma) = 2^{-8\tau} C^{-2} \sigma^{2\tau} \eta$  attains its maximum with respect to  $\eta < s$  for  $\eta = s$ . Besides, under the constraint  $s + \sigma = S$  the function  $\epsilon(s, \sigma)$  attains its maximum when  $\sigma = 2\tau s$  and  $s = \frac{S}{1+2\tau}$ . Hence,  $S$  being fixed, the domain of  $\psi$  contains  $\epsilon B_S^F$  if

$$\epsilon < 2^{-8\tau} C^{-2} \frac{S}{1+2\tau} \left( \frac{2\tau S}{12(1+2\tau)} \right)^{2\tau}.$$

Given that  $S < 1 < \tau$  by hypothesis, it suffices that  $\epsilon$  be equal to the stated value.  $\square$

**A.1. Regularity of the right-inverse.** In the proof of theorem 8 we have built right inverses  $\psi : \epsilon B_{s+\eta+\sigma}^F \rightarrow \eta B_{s+\eta}^E$ , of  $\phi$ , commuting with inclusions. The estimate given in the statement shows that  $\psi$  is continuous at 0; due to the invariance of the hypotheses of the theorem by small translations,  $\psi$  is locally continuous.

We further make the following two assumptions:

- The maps  $\phi'(x)^{-1} : F_{s+\sigma} \rightarrow E_s$  are left (as well as right) inverses (in theorem 4 we have restricted to an adequate class of symplectomorphisms);
- The scale  $(|\cdot|_s)$  of norms of  $(E_s)$  satisfies some interpolation inequality:

$$|x|_{s+\sigma}^2 \leq |x|_s |x|_{s+\tilde{\sigma}} \quad \text{for all } s, \sigma, \tilde{\sigma} = \sigma \left( 1 + \frac{1}{s} \right)$$

(according to the remark after corollary 16, this estimate is satisfied in the case of interest to us, since  $\sigma + \log(1 + \sigma/s) \leq \tilde{\sigma}$ ).

**Lemma 11** (Lipschitz regularity). *If  $\sigma < s$  and  $y, \hat{y} \in \epsilon B_{s+\sigma}^F$  with  $\epsilon = 2^{-14\tau} C^{-3} \sigma^{3\tau}$ ,*

$$|\psi(\hat{y}) - \psi(y)|_s \leq C_L |\hat{y} - y|_{s+\sigma}, \quad C_L = 2C' \sigma^{-\tau'}.$$

*In particular,  $\psi$  is the unique local right inverse of  $\phi$ , and hence is also its local left inverse.*

*Proof.* Fix  $\eta < \zeta < \sigma < s$ ; the impatient reader can readily look at the end of the proof how to choose the auxiliary parameters  $\eta$  and  $\zeta$  more precisely.

Let  $\epsilon = 2^{-8\tau} C^{-2} \zeta^{2\tau} \eta$ , and  $y, \hat{y} \in \epsilon B_{s+\sigma}^F$ . According to theorem 8,  $x := \psi(y)$  and  $\hat{x} := \psi(\hat{y})$  are in  $\eta B_{s+\sigma-\zeta}^E$ , provided the condition, to be checked later, that  $\eta < s + \sigma - \zeta$ . In particular, we will use a priori that

$$|\hat{x} - x|_{s+\sigma-\zeta} \leq |\hat{x}|_{s+\sigma-\zeta} + |x|_{s+\sigma-\zeta} \leq 2\eta.$$

We have

$$\begin{aligned} \hat{x} - x &= \phi'(x)^{-1} \phi'(x)(\hat{x} - x) \\ &= \phi'(x)^{-1} (\hat{y} - y - Q(x, \hat{x})) \end{aligned}$$

and, according to the assumed estimate on  $\phi'(x)^{-1}$  and to lemma 9,

$$|\hat{x} - x|_s \leq C' \sigma^{-\tau'} |\hat{y} - y|_{s+\sigma} + 2^{-1} C \zeta^{-\tau} |\hat{x} - x|_{s+2\eta+|\hat{x}-x|_s}^2.$$

In the norm index of the last term, we will coarsely bound  $|\hat{x} - x|_s$  by  $2\eta$ . Additionally using the interpolation inequality:

$$|\hat{x} - x|_{s+4\eta}^2 \leq |\hat{x} - x|_s |\hat{x} - x|_{s+\tilde{\sigma}}, \quad \tilde{\sigma} = 4\eta \left(1 + \frac{1}{s}\right),$$

yields

$$(1 - 2^{-1} C \zeta^{-\tau} |\hat{x} - x|_{s+\tilde{\sigma}}) |\hat{x} - x|_s \leq C' \sigma^{-\tau'} |\hat{y} - y|_{s+\sigma}.$$

Now, we want to choose  $\eta$  small enough so that

– first,  $\tilde{\sigma} \leq \sigma - \zeta$ , which implies  $|\hat{x} - x|_{s+\tilde{\sigma}} \leq 2\eta$ . By definition of  $\tilde{\sigma}$ , it suffices to have  $\eta \leq \frac{\sigma - \zeta}{4(1+1/s)}$ .

– second,  $2^{-1} C \zeta^{-\tau} 2\eta \leq 1/2$ , or  $\eta \leq \frac{\zeta^\tau}{2C}$ , which implies that  $2^{-1} C \zeta^{-\tau} |\hat{x} - x|_{s+\tilde{\sigma}} \leq 1/2$ , and hence  $|\hat{x} - x|_s \leq 2C' \sigma^{-\tau'} |\hat{y} - y|_{s+\sigma}$ .

A choice is  $\zeta = \frac{\sigma}{2}$  and  $\eta = \frac{\sigma^\tau}{16C} < s$ , whence the value of  $\epsilon$  in the statement.  $\square$

**Proposition 12** (Smoothness). *For every  $\sigma < s$ , there exists  $\epsilon, C_1$  such that for every  $y, \hat{y} \in \epsilon B_{s+\sigma}^F$ ,*

$$|\psi(\hat{y}) - \psi(y) - \phi'(\psi(y))^{-1}(\hat{y} - y)|_s \leq C_1 |\hat{y} - y|_{s+\sigma}^2.$$

*Moreover, the map  $\psi' : \epsilon B_{s+\sigma}^F \rightarrow L(F_{s+\sigma}, E_s)$  defined locally by  $\psi'(y) = \phi'(\psi(y))^{-1}$  is continuous.*

*Proof.* Fix  $\epsilon$  as in the previous proof and  $y, \hat{y} \in \varepsilon B_{s+\sigma}^F$ . Let  $x = \psi(y)$ ,  $\eta = \hat{y} - y$ ,  $\xi = \psi(y + \eta) - \psi(y)$  (thus  $\eta = \phi(x + \xi) - \phi(x)$ ), and  $\Delta := \psi(y + \eta) - \psi(y) - \phi'(x)^{-1}\eta$ . Definitions yield

$$\Delta = \phi'(x)^{-1} (\phi'(x)\xi - \eta) = -\phi'(x)^{-1}Q(x, x + \xi).$$

Using the estimates on  $\phi'(x)^{-1}$  and  $Q$  and the latter lemma,

$$|\Delta|_s \leq C_1 |\eta|_{s+\sigma'}^2$$

for some  $\sigma'$  tending to 0 when  $\sigma$  itself tends to 0, and for some  $C_1 > 0$  depending on  $\sigma$ . Up the substitution of  $\sigma$  by  $\sigma'$ , the estimate is proved.

The inversion of linear operators between Banach spaces being analytic,  $y \mapsto \phi(\psi(y))^{-1}$  is continuous in the stated sense.  $\square$

**Corollary 13.** *If  $\pi \in L(E_s, V)$  is a family of linear maps, commuting with inclusions, into a fixed Banach space  $V$ , then  $\pi \circ \psi$  is  $C^1$  and  $(\pi \circ \psi)' = \pi \cdot \phi' \circ \psi$ .*

This corollary is used with  $\pi : (K, G, \beta) \mapsto \beta$  in the proof of theorem 4.

## B. SOME ESTIMATES ON ANALYTIC ISOMORPHISMS

In this appendix, we give a quantitative inverse function theorem for real analytic isomorphisms on  $\mathbb{T}_s^n$ . This is used in section 2, to parametrize locally  $\mathcal{D}_s$  by vector fields, and, in lemma 5, to solve the cohomological equation for the frequency offset  $\delta\beta$ .

Recall that we have set  $\mathbb{T}_s^n := \{\theta \in \mathbb{C}^n / 2\pi\mathbb{Z}^n, \max_{1 \leq j \leq n} |\operatorname{Im} \theta_j| \leq s\}$ . We will denote by  $p : \mathbb{R}_s^n := \mathbb{R}^n \times i[-s, s]^n \rightarrow \mathbb{T}_s^n$  its universal covering.

**Proposition 14.** *Let  $v \in \mathcal{A}(\mathbb{T}_{s+2\sigma}^n, \mathbb{C}^n)$ ,  $|v|_{s+2\sigma} < \sigma$ . The map  $\operatorname{id} + v : \mathbb{T}_{s+2\sigma}^n \rightarrow \mathbb{R}_{s+3\sigma}^n$  induces a map  $\varphi : \mathbb{T}_{s+2\sigma}^n \rightarrow \mathbb{T}_{s+3\sigma}^n$  whose restriction  $\varphi : \mathbb{T}_{s+\sigma}^n \rightarrow \mathbb{T}_{s+2\sigma}^n$  has a unique right inverse  $\psi : \mathbb{T}_s^n \rightarrow \mathbb{T}_{s+\sigma}^n$ :*

$$\begin{array}{ccc} \mathbb{T}_{s+\sigma}^n & \xrightarrow{\varphi} & \mathbb{T}_{s+2\sigma}^n \\ & \nwarrow \psi & \uparrow \\ & & \mathbb{T}_s^n \end{array}$$

Furthermore,

$$|\psi - \operatorname{id}|_s \leq |v|_{s+\sigma}$$

and, provided  $2\sigma^{-1}|v|_{s+2\sigma} \leq 1$ ,

$$|\psi' - \operatorname{id}| \leq 2\sigma^{-1}|v|_{s+2\sigma}.$$

*Proof.* Let  $\Phi : \mathbb{R}_{s+2\sigma}^n \rightarrow \mathbb{R}_{s+3\sigma}^n$  be a continuous lift of  $\operatorname{id} + v$  and  $k \in M_n(\mathbb{Z})$ ,  $k(l) := \Phi(x + l) - \Phi(x)$ .

- (1) *Injectivity of  $\Phi : \mathbb{R}_{s+\sigma}^n \rightarrow \mathbb{R}_{s+2\sigma}^n$ .* Suppose that  $x, \hat{x} \in \mathbb{R}_{s+\sigma}^n$  and  $\Phi(x) = \Phi(\hat{x})$ . By the mean value theorem,

$$|x - \hat{x}| = |v(p\hat{x}) - v(px)| \leq |v'|_{s+\sigma}|x - \hat{x}|,$$

and, by Cauchy's inequality,

$$|x - \hat{x}| \leq \frac{|v|_{s+2\sigma}}{\sigma} |x - \hat{x}| < |\hat{x} - x|,$$

hence  $x = \hat{x}$ .

- (2) *Surjectivity of  $\Phi$* :  $\mathbb{R}_s^n \subset \Phi(\mathbb{R}_{s+\sigma}^n)$ . For any given  $y \in \mathbb{R}_s^n$ , the contraction

$$f : \mathbb{R}_{s+\sigma}^n \rightarrow \mathbb{R}_{s+\sigma}^n, \quad x \mapsto y - v(x)$$

has a unique fixed point, which is a pre-image of  $y$  by  $\Phi$ .

- (3) *Injectivity of  $\varphi$* :  $\mathbb{T}_{s+\sigma}^n \rightarrow \mathbb{T}_{s+2\sigma}^n$ . Suppose that  $px, p\hat{x} \in \mathbb{R}_{s+\sigma}^n$  and  $\varphi(px) = \varphi(p\hat{x})$ , i.e.  $\Phi(x) = \Phi(\hat{x}) + \kappa$  for some  $\kappa \in \mathbb{Z}^n$ . That  $k$  be in  $GL(n, \mathbb{Z})$ , follows from the invertibility of  $\Phi$ . Hence,  $\Phi(x - k^{-1}(\kappa)) = \Phi(\hat{x})$ , and, due to the injectivity of  $\Phi$ ,  $px = p\hat{x}$ .
- (4) *Surjectivity of  $\varphi$* :  $\mathbb{T}_s^n \subset \varphi(\mathbb{T}_{s+\sigma}^n)$ . This is a trivial consequence of that of  $\Phi$ .
- (5) *Estimate on  $\psi$* :  $\psi := \varphi^{-1} : \mathbb{T}_s^n \rightarrow \mathbb{T}_{s+\sigma}^n$ . Note that the wanted estimate on  $\psi$  is in the sense of  $\Psi := \Phi^{-1} : \mathbb{R}_s^n \rightarrow \mathbb{R}_{s+\sigma}^n$ . If  $y \in \mathbb{R}_s^n$ ,

$$\Psi(y) - y = -v(p\Psi(y)),$$

hence  $|\Psi - \text{id}|_s \leq |v|_{s+\sigma}$ .

- (6) *Estimate on  $\psi'$* . We have  $\psi' = \varphi'^{-1} \circ \varphi$ , where  $\varphi'^{-1}(x)$  stands for the inverse of the map  $\xi \mapsto \varphi'(x) \cdot \xi$ . Hence

$$\psi' - \text{id} = \varphi'^{-1} \circ \varphi - \text{id},$$

and, under the assumption that  $2\sigma^{-1}|v|_{s+2\sigma} \leq 1$ ,

$$|\psi' - \text{id}|_s \leq |\varphi'^{-1} - \text{id}|_{s+\sigma} \leq \frac{|v'|_{s+\sigma}}{1 - |v'|_{s+\sigma}} \leq \frac{\sigma^{-1}|v|_{s+2\sigma}}{1 - \sigma^{-1}|v|_{s+2\sigma}} \leq 2\sigma^{-1}|v|_{s+2\sigma}.$$

□

### C. INTERPOLATION OF SPACES OF ANALYTIC FUNCTIONS

In this section we prove some Hadamard interpolation inequalities, which are used in A.1.

Recall that we denote by  $\mathbb{T}_{\mathbb{C}}^n$  the infinite annulus  $\mathbb{C}^n/2\pi\mathbb{Z}^n$ , by  $\mathbb{T}_s^n$ ,  $s > 0$ , the bounded sub-annulus  $\{\theta \in \mathbb{T}_{\mathbb{C}}^n, |\text{Im } \theta_j| \leq s, j = 1 \dots n\}$  and by  $\mathbb{D}_t^n$ ,  $t > 0$ , the polydisc  $\{r \in \mathbb{C}^n, |r_j| \leq t, j = 1 \dots n\}$ . The supremum norm of a function  $f \in \mathcal{A}(\mathbb{T}_s^n \times \mathbb{D}_t^n)$  will be denoted by  $|f|_{s,t}$ .

Let  $0 < s_0 \leq s_1$  and  $0 < t_0 \leq t_1$  be such that

$$\log \frac{t_1}{t_0} = s_1 - s_0.$$

Let also  $0 \leq \rho \leq 1$  and

$$s = (1 - \rho)s_0 + \rho s_1 \quad \text{and} \quad t = t_0^{1-\rho} t_1^\rho.$$

**Proposition 15.** *If  $f \in \mathcal{A}(\mathbb{T}_{s_1}^n \times \mathbb{D}_{t_1}^n)$ ,*

$$|f|_{s,t} \leq |f|_{s_0,t_0}^{1-\rho} |f|_{s_1,t_1}^\rho.$$

*Proof.* Let  $\tilde{f}$  be the function on  $\mathbb{T}_{s_1}^n \times \mathbb{D}_{t_1}^n$ , constant on  $2n$ -tori of equations  $(\operatorname{Im} \theta, r) = cst$ , defined by

$$\tilde{f}(\theta, r) = \max_{\mu, \nu \in \mathbb{T}^n} |f((\pm\theta_1 + \mu_1, \dots, \pm\theta_n + \mu_n), (r_1 e^{i\nu_1}, \dots, r_n e^{i\nu_n}))|$$

(with all possible combinations of signs). Since  $\log |f|$  is subharmonic and  $\mathbb{T}^{2n}$  is compact,  $\log \tilde{f}$  too is upper semi-continuous. Besides,  $\log \tilde{f}$  satisfies the mean inequality, hence is plurisubharmonic.

By the maximum principle, the restriction of  $|f|$  to  $\mathbb{T}_s^n \times \mathbb{D}_t^n$  attains its maximum on the distinguished boundary of  $\mathbb{T}_s^n \times \mathbb{D}_t^n$ . Due to the symmetry of  $\tilde{f}$ :

$$|f|_{s,t} = \tilde{f}(is\epsilon, t\epsilon), \quad \epsilon = (1, \dots, 1).$$

Now, the function

$$\varphi(z) := \tilde{f}(z\epsilon, e^{-(iz+s)}t\epsilon)$$

is well defined on  $\mathbb{T}_{s_1}$ , for it is constant with respect to  $\operatorname{Re} z$  and, due to the relations imposed on the norm indices, if  $|\operatorname{Im} z| \leq s_1$  then  $|e^{-(iz+s)}t| \leq e^{s_1-s}t = t_1$ .

The estimate

$$\log \varphi(z) \leq \frac{s_1 - \operatorname{Im} z}{s_1 - s_0} \varphi(s_0 i) + \frac{\operatorname{Im} z - s_0}{s_1 - s_0} \varphi(s_1 i)$$

trivially holds if  $\operatorname{Im} z = s_0$  or  $s_1$ , for, as noted above for  $j = 1$ ,  $e^{s_j-s}t = t_j$ ,  $j = 0, 1$ . But note that the left and right hand sides respectively are subharmonic and harmonic. Hence the estimate holds whenever  $s_0 \leq \operatorname{Im} z \leq s_1$ , whence the claim for  $z = is$ .  $\square$

Recall that we have let  $\mathbb{T}_s^n := \mathbb{T}_s^n \times \mathbb{D}_s^n$ ,  $s > 0$ , and, for a function  $f \in \mathcal{A}(\mathbb{T}_s^n)$ , let  $|f|_s = |f|_{s,s}$  denote its supremum norm on  $\mathbb{T}_s^n$ . As in the rest of the paper, we now restrict the discussion to widths of analyticity  $\leq 1$ .

**Corollary 16.** *If  $\sigma_1 = -\log(1 - \frac{\sigma_0}{s})$  and  $f \in \mathcal{A}(\mathbb{T}_{s+\sigma_1}^n)$ ,*

$$|f|_s^2 \leq |f|_{s-\sigma_0} |f|_{s+\sigma_1}.$$

In A.1, we will use the equivalent fact that, if  $\tilde{\sigma} = s + \log(1 + \frac{\sigma}{s})$  and  $f \in \mathcal{A}(\mathbb{T}_{s+\tilde{\sigma}}^n)$ ,

$$|f|_{s+\sigma}^2 \leq |f|_s |f|_{s+\tilde{\sigma}}.$$

*Proof.* In proposition 15, consider the following particular case :

- $\rho = 1/2$ . Hence

$$s = \frac{s_0 + s_1}{2} \quad \text{and} \quad t = \sqrt{t_0 t_1}.$$

- $s = t$ . Hence in particular  $t_0 = s e^{s_0-s}$  and  $t_1 = s e^{s_1-s}$ .

Then

$$|f|_s^2 = |f|_{s,s}^2 \leq |f|_{s_0,t_0} |f|_{s_1,t_1}.$$

We want to determine  $\max(s_0, t_0)$  and  $\max(s_1, t_1)$ . Let  $\sigma_1 := s - s_0 = s_1 - s$ . Then  $t_0 = s e^{-\sigma_1}$  and  $t_1 = s e^{\sigma_1}$ . The expression  $s + \sigma - s e^\sigma$  has the sign of  $\sigma$  (in the relevant

region  $0 \leq s + \sigma \leq 1$ ,  $0 \leq s \leq 1$ ); by evaluating it at  $\sigma = \pm\sigma_1$ , we see that  $s_0 \leq t_0$  and  $s_1 \geq t_1$ .

Therefore, since the norm  $|\cdot|_{s,t}$  is non-decreasing with respect to both  $s$  and  $t$ ,

$$|f|_s^2 \leq |f|_{t_0,t_0} |f|_{s_1,s_1} = |f|_{t_0} |f|_{s_1}$$

(thus giving up estimates uniform with respect to small values of  $s$ ). By further setting  $\sigma_0 = s - t_0 = s(1 - e^{-\sigma_1})$ , we get the wanted estimate, and the asserted relation between  $\sigma_0$  and  $\sigma_1$  is readily verified.  $\square$

#### D. WEAKER ARITHMETIC CONDITIONS OF CONVERGENCE

In this section, we look more carefully to the arithmetic conditions needed for the induction to converge, in the proof of the inverse function theorem 8.

A function  $\Delta : \mathbb{N}_* \rightarrow [1, +\infty[$  being given, define the set  $D_\Delta$  as the subset of vectors  $\alpha \in \mathbb{R}^n$  such that

$$|k \cdot \alpha| \geq \frac{(|k| + n - 1)^{n-1}}{\Delta(|k|)} \quad (\forall k \in \mathbb{Z}^n \setminus \{0\}).$$

(The function  $\Delta$  is just some other normalization of what is an approximation function in Rüssmann [1975] or a zone function in Dumas et al. [2004].) For  $D_\Delta$  to be non empty, trivially we need  $\lim_{+\infty} \Delta = +\infty$ .

**Proposition 17.** *The conclusions of theorems 4 and 1 hold if there exist  $c > 0$  and  $\delta \in ]0, 1[$  such that*

$$\sum_{\ell \geq 1} \Delta(\ell) e^{-\ell/j^2} \leq \exp(c 2^{\delta j}) \quad \text{as } j \rightarrow +\infty.$$

**Example 18** The Diophantine set  $D_{\gamma,\tau}$  corresponds to a polynomially growing function  $\Delta$ , and to a polynomially growing function  $\sum_{\ell \geq 1} \Delta(\ell) e^{-\ell 2^{-j}}$ . A foriori,  $\sum_{\ell \geq 1} \Delta(\ell) e^{-\ell/j^2}$  is at most polynomially growing.

*Proof.* Call  $L$  the discrete Laplace transform of  $\Delta$ :

$$L(\sigma) = \sum_{\ell \geq 1} \Delta(\ell) e^{-\ell\sigma},$$

and assume it is finite for all  $\sigma > 0$ . Patterning the proof of lemma 5, we get the following generalization.

**Lemma 19.** *Let  $g \in \mathcal{A}(\mathbb{T}_{s+\sigma}^n)$  having 0-average. There is a unique function  $f \in \mathcal{A}(\mathbb{T}_s^n)$  of zero average such that  $\mathcal{L}_\alpha f = g$ . This function satisfies*

$$|f|_s \leq C L(\sigma) |g|_{s+\sigma}, \quad C = \frac{2^n e}{(n-1)!}.$$

(Again, see Rüssmann [1975] for improved estimates. But such an improvement is not the crux of our purpose here.)

Taking up the proof of the inverse function theorem of appendix A with our new estimates (see in particular equation (2)), we see that the Newton algorithm converges provided

$$\sum_{j \geq 0} 2^{-j} \log L(\sigma_j) < \infty,$$

for some choice of the converging series  $\sum \sigma_j$ . Choosing  $\sum \sigma_j = \sum j^{-2}$ , we see that it is enough that  $\log L(\sigma_j) \leq c 2^{\delta j}$  for some  $c > 0$  and  $\delta \in ]0, 1[$ , whence the given criterion.  $\square$

## E. COMMENTS

*Section 1.* The proof of Kolmogorov's theorem presented here differs from others chiefly for the following reasons:

- The seeming detour through Herman's normal form reduces Kolmogorov's theorem to a functionally well posed inversion problem (compare with Zehnder [1975, 1976]). This powerful trick consists in switching the frequency obstruction (obstruction to the conjugacy to the initial dynamics) from one side of the conjugacy to the other. It was extensively used in Moser [1967]. The remaining, finite dimensional problem is then to show that the frequency offset  $\beta \in \mathbb{R}^n$  may vanish; in general, it is met using a non-degeneracy hypothesis of one kind or another. Looking backward, this last step is not the most difficult, but was probably not well understood before M. Herman in the 80s (see Rüssmann [1990] and Sevryuk [1999]). The functional setting chosen here adapts to more degenerate cases, including lower dimensional tori, in a straightforward manner (see Féjoz [2004]; compare to Herman's preferred proof for Lagrangian tori, as exposed in Bost [1986]).
- Classical perturbation series (or some modification of these) have been shown to converge in some cases (Siegel [1942] for the convergence of Schröder series in the Siegel problem, see Eliasson [1996] for Lindstedt series of Hamiltonians). Direct methods for proving their convergence are involved because, as J. Moser noticed in [Moser, 1967, p. 149], these series do not converge absolutely, and thus the proof of semi-convergence must take into account compensations or the precise accumulation of small denominators through a subtle combinatorial analysis. On the other hand, the perturbation series yielded by the Newton algorithm are absolutely convergent, provided that one adequately chooses the width of analytic spaces at each step of the induction. This was a major discovery of Kolmogorov. In the first approximation, the series so obtained can be thought of as obtained by grouping terms of the classical perturbation series (from step  $j$  to step  $j + 1$ , the non resonant terms of size  $\epsilon^{2^j}, \dots, \epsilon^{2^{j+1}-1}$  are eliminated). The magic is that compensations are taken into account without noticing, and it would be interesting to understand how classical and Newton series relate precisely, maybe with mould calculus.
- We encapsulate the Newton algorithm in an abstract inverse function theorem à la Nash-Moser. The algorithm indeed converges without any specific hypothesis on the internal structure of the variables. At the expense of some optimality, ignoring this structure allows for simple estimates (and control of the bounds) and for solving a whole class of analogous problems with the same toolbox (lower dimensional tori, codimension-one tori, Siegel problem, as well as some problems in singularity theory).



- The analytic (or Gevrey) category is simpler, in Nash-Moser theory, than Hölder or Sobolev categories because the Newton algorithm can be carried out without intercalating smoothing operators (cf. Sergeraert [1972], Bost [1986]).
- Incidentally, Hadamard interpolation inequalities are simple to infer for analytic norms because, again, they do not depend on regularizing operators, as it is shown in appendix C (cf. [Hörmander, 1976, Theorem A.5]).
- The use of auxiliary norms ( $|\cdot|_{G,s}$  in lemmas 5 and 7,  $|\cdot|_{x,s}$  in appendix A) prevents from artificially loosing, due to compositions, a fixed width of analyticity at each step of the Newton algorithm –the domains of analyticity being deformed rather than shrunk. As a pitfall, the argument of [Jacobowitz, 1972, Sections 5 and 6] to deduce an analytic function theorem in the smooth category abstractly from the theorem in the analytic category, does not apply directly here (see comment below).

*Section 1. Theorem 4.* Herman’s normal form is the Hamiltonian analogue of the normal form of vector fields on the torus in the neighborhood of Diophantine constant vector fields (Arnold [1961], Moser [1966a]). The normal form for Hamiltonians implies the normal form for vector fields on the torus [Féjoz, 2004, Théorème 40] and is actually simpler to prove from the algebraic point of view.

*Section 3. Lemma 5.* The estimate is obtained by bounding the terms of Fourier series one by one. In a more careful estimate, one should take into account the fact that if  $|k \cdot \alpha|$  is small, then  $k' \cdot \alpha$  is not so small for neighboring  $k'$ ’s. This allows to find the optimal exponent of  $\sigma$ , making it independant of the dimension; see Moser [1966b] and Rüssmann [1975].

*Appendix A. Theorem 8.* – The two competing small parameters  $\eta$  and  $\sigma$  being fixed, our choice of the sequence  $(\sigma_n)$  maximizes  $\epsilon$  for the Newton algorithm. It does not modify the sequence  $(x_k)$  but only the information we retain from  $(x_k)$ .

- In the expression of  $\epsilon$ , the square exponent of  $C$  is inherent in the quadratic convergence of Newton’s algorithm. From this follows the dependance, in KAM theory, of the size  $\epsilon$  of the allowed perturbation with respect to the small diophantine constant  $\gamma$ :  $\epsilon = O(\gamma^2)$ .
- The method of Jacobowitz [1972] (see Moser [1966b] also in order to deduce an inverse function theorem in the smooth category from its analogue in the analytic category does not work directly, here. The idea would be to use Jackson’s theorem in approximation theory to characterize the Hölder spaces by their approximation properties in terms of analytic functions and, then, to find a smooth preimage  $x$  by  $\phi$  of a smooth function  $y$  as the limit of analytic preimages  $x_j$  of analytic approximations  $y_j$  of  $y$ . However, in our inversion function theorem we require the operator  $\phi$  to be defined only on balls  $\sigma B_{s+\sigma}$  with shrinking radii when  $s+\sigma$  tends to 0. This domain is too small in general to include all the analytic approximations  $y_j$  of a smooth  $y$ . Such a restriction is inherent in the presence of composition operators. Jacobowitz [1972] did not have to deal with such operators for the problem of isometric embeddings. Yet we *could* generalize Jacobowitz’s proof at the expense of making additionnal hypotheses on the form of our operator  $\phi$ ,

which would take into account the specificity of directions  $K$  and  $G$ , as well as of the real phase space and of its complex extension.

*Appendix A.1.* It is possible to prove that  $\psi$  is  $C^1$  without additional assumptions, just by patterning [Sergeraert, 1972, p. 626]). Yet the proof simplifies and the estimates improve under the combined two additional assumptions. In particular, the existence of a right inverse of  $\phi'(x)$  makes the inverse  $\psi$  unique and thus allows to ignore the way it was built.

*Appendix B.* We include this elementary section for the sake of completeness, although the quantitative estimates are needed only if one wants a quantitative version of Kolmogorov's theorem, with an explicit value of  $\epsilon$ . A similar proposition (for germs at a point of maps in  $\mathbb{C}^n$ ) is proved in Pöschel [2001] using a more sophisticated argument from degree theory.

*Appendix C.* In this paragraph, the obtained inequalities generalize the standard Hadamard convexity inequalities. They are optimal and show that analytic norms are not quite convex with respect to the width of the complex extensions, due to the geometry of the phase space. See [Narasimhan, 1995, Chap. 8] for more general but less precise inequalities.

*Appendix E. Proposition 17.* There are reasons to believe that the so obtained arithmetic condition is not optimal. Indeed, solving the exact cohomological equation at each step is inefficient because the small denominators appearing with intermediate-order harmonics deteriorate the estimates, whereas some of these harmonics could have a smaller amplitude than the error terms and thus would better not be taken care of. Even stronger, Rüssmann and Pöschel remarkably and recently noticed that at each step it is worth neglecting part of the low-order harmonics themselves (to some carefully chosen extent). Then the expense, a worse error term, turns out to be cheaper than that the gain –namely, the right hand side of the cohomological equation now has a smaller size over a larger complex extension. This allows, with a slowly converging sequence of approximations, to show the persistence of invariant tori under some arithmetic condition which, in one dimension, is equivalent to the Brjuno condition; see Pöschel [2009].

*Thank you to P. Bernard, A. Chenciner, R. Krikorian, I. Kupka, D. Sauzin and J.-C. Yoccoz, for illuminating discussions, and to A. Albouy and A. Knauf for careful reading and correcting.*

## REFERENCES

- V. I. Arnold. Small denominators. I. Mapping the circle onto itself. *Izv. Akad. Nauk SSSR Ser. Mat.*, 25:21–86, 1961. ISSN 0373-2436.
- V. I. Arnold. Small denominators and problems of stability of motion in classical and celestial mechanics. *Uspehi Mat. Nauk*, 18(6 (114)):91–192, 1963. ISSN 0042-1316.
- V. I. Arnold. Instability of dynamical systems with many degrees of freedom. *Dokl. Akad. Nauk SSSR*, 156:9–12, 1964. ISSN 0002-3264.

- J.-B. Bost. Tores invariants des systèmes dynamiques hamiltoniens (d'après Kolmogorov, Arnold, Moser, Rüssmann, Zehnder, Herman, Pöschel, ...). *Astérisque*, 1984/85(133-134):113–157, 1986. ISSN 0303-1179. Séminaire Bourbaki.
- A. Celletti and L. Chierchia. KAM stability and celestial mechanics. *Mem. Amer. Math. Soc.*, 187(878):viii+134, 2007. ISSN 0065-9266.
- L. Chierchia. A. N. Kolmogorov's 1954 paper on nearly-integrable Hamiltonian systems. A comment on: "On conservation of conditionally periodic motions for a small change in Hamilton's function" [Dokl. Akad. Nauk SSSR (N.S.) **98** (1954), 527–530; mr0068687]. *Regul. Chaotic Dyn.*, 13(2):130–139, 2008. ISSN 1560-3547.
- R. de la Llave. A tutorial on KAM theory. In *Smooth ergodic theory and its applications (Seattle, WA, 1999)*, volume 69 of *Proc. Sympos. Pure Math.*, pages 175–292. Amer. Math. Soc., Providence, RI, 2001.
- H. S. Dumas, J. A. Ellison, and M. Vogt. First-order averaging principles for maps with applications to accelerator beam dynamics. *SIAM J. Appl. Dyn. Syst.*, 3(4):409–432 (electronic), 2004. ISSN 1536-0040. doi: 10.1137/030600436. URL <http://dx.doi.org/10.1137/030600436>.
- L. H. Eliasson. Absolutely convergent series expansions for quasi periodic motions. *Math. Phys. Electron. J.*, 2:Paper 4, 33 pp. (electronic), 1996. ISSN 1086-6655.
- J. Féjoz. Démonstration du 'théorème d'Arnold' sur la stabilité du système planétaire (d'après Herman). *Ergodic Theory Dynam. Systems*, 24(5):1521–1582, 2004. ISSN 0143-3857. doi: 10.1017/S0143385704000410. URL <http://dx.doi.org/10.1017/S0143385704000410>.
- L. Hörmander. The boundary problems of physical geodesy. *Arch. Rational Mech. Anal.*, 62(1):1–52, 1976. ISSN 0003-9527.
- H. Jacobowitz. Implicit function theorems and isometric embeddings. *Ann. of Math. (2)*, 95:191–225, 1972. ISSN 0003-486X.
- A. N. Kolmogorov. On conservation of conditionally periodic motions for a small change in Hamilton's function. *Dokl. Akad. Nauk SSSR (N.S.)*, 98:527–530, 1954. ISSN 0002-3264.
- J. Moser. A rapidly convergent iteration method and non-linear partial differential equations. I. *Ann. Scuola Norm. Sup. Pisa (3)*, 20:265–315, 1966a.
- J. Moser. A rapidly convergent iteration method and non-linear differential equations. II. *Ann. Scuola Norm. Sup. Pisa (3)*, 20:499–535, 1966b.
- J. Moser. Convergent series expansions for quasi-periodic motions. *Math. Ann.*, 169:136–176, 1967. ISSN 0025-5831.
- R. Narasimhan. *Several complex variables*. Chicago Lectures in Mathematics. University of Chicago Press, Chicago, IL, 1995. ISBN 0-226-56817-2. Reprint of the 1971 original.
- J. Pöschel. A lecture on the classical KAM theorem. In *Smooth ergodic theory and its applications (Seattle, WA, 1999)*, volume 69 of *Proc. Sympos. Pure Math.*, pages 707–732. Amer. Math. Soc., Providence, RI, 2001.
- J. Pöschel. A second lecture on the classical KAM theorem – Rüssmann's scheme. Preprint, 2009.
- H. Rüssmann. On optimal estimates for the solutions of linear partial differential equations of first order with constant coefficients on the torus. In *Dynamical systems, theory and applications (Rencontres, Battelle Res. Inst., Seattle, Wash., 1974)*, pages

- 598–624. Lecture Notes in Phys., Vol. 38. Springer, Berlin, 1975.
- H. Rüssmann. Nondegeneracy in the perturbation theory of integrable dynamical systems. In *Stochastics, algebra and analysis in classical and quantum dynamics (Marseille, 1988)*, volume 59 of *Math. Appl.*, pages 211–223. Kluwer Acad. Publ., Dordrecht, 1990.
- F. Sergeraert. Un théorème de fonctions implicites sur certains espaces de Fréchet et quelques applications. *Ann. Sci. École Norm. Sup. (4)*, 5:599–660, 1972. ISSN 0012-9593.
- M. B. Sevryuk. The lack-of-parameters problem in the KAM theory revisited. In *Hamiltonian systems with three or more degrees of freedom (S’Agaró, 1995)*, volume 533 of *NATO Adv. Sci. Inst. Ser. C Math. Phys. Sci.*, pages 568–572. Kluwer Acad. Publ., Dordrecht, 1999.
- M. B. Sevryuk. The classical KAM theory at the dawn of the twenty-first century. *Mosc. Math. J.*, 3(3):1113–1144, 1201–1202, 2003. ISSN 1609-3321. {Dedicated to Vladimir Igorevich Arnold on the occasion of his 65th birthday}.
- C. L. Siegel. Iteration of analytic functions. *Ann. of Math. (2)*, 43:607–612, 1942. ISSN 0003-486X.
- E. Zehnder. Generalized implicit function theorems with applications to some small divisor problems. I. *Comm. Pure Appl. Math.*, 28:91–140, 1975. ISSN 0010-3640.
- E. Zehnder. Generalized implicit function theorems with applications to some small divisor problems. II. *Comm. Pure Appl. Math.*, 29(1):49–111, 1976. ISSN 0010-3640.